Universitat Oberta de Catalunya (UOC)

Máster Universitario en Ciencia de Datos (*Data Science*)

TRABAJO FINAL DE MÁSTER

Área: Machine Learning

Analysis of Political Ideology on Social Media and Official Platforms

Spain and the United Kingdom

Autor: Victor García Domingo

Tutor: Joaquim Moré López

Profesor: Jordi Conesa Caralt

Barcelona, January 3, 2021

Credits/Copyright

FICHA DEL TRABAJO FINAL

Título del trabajo:	Analysis of Political Ideology on Social Media and Official Platforms: Spain and the United Kingdom
Nombre del autor:	Victor García Domingo
Nombre del colaborador/a docente:	Joaquim Moré López
Nombre del PRA:	Jordi Conesa Caralt
Fecha de entrega (mm/aaaa):	01/2021
Titulación o programa:	Máster en Ciencia de Datos
Área del Trabajo Final:	Machine Learning
Idioma del trabajo:	inglés
Palabras clave:	NLP, political ideology, twitter

Dedication

To my parents, for giving me everything and more.

To Eric, for being the best brother I could have.

To Ricard, for motivating me to study Computer Science.
Get well soon.

To Jalpa, a fundamental pillar of my life.

"*Stayed in Mississippi a day too long.*"

— BOB DYLAN

Acknowledgements

I would like to express my special thanks to Joaquim Moré for his expert guidance.

To Jalpa Soni, for supporting me during this project.

To my friends, for not giving up on me during the last five years.

Abstract

Durante las campañas electorales, los partidos políticos usan varios medios para hacer llegar al ciudadano mensajes para influir en su voto, siendo dos de los más importantes las redes sociales y los programas electorales.

Este trabajo tiene como objetivo analizar y comparar la orientación ideológica de los mensajes publicados por los partidos políticos según el canal utilizado y a partir de diferentes temas entre los que se muestran los resultados de *igualdad*.

Para ello, se usará como corpus una selección de tuits y programas electorales de los cinco partidos políticos nacionales con más representación en las últimas elecciones de España y cuatro de las últimas elecciones en el Reino Unido.

El estudio permitirá observar si el mensaje político en los tweets se radicaliza respecto al mensaje de los programas electorales para cada uno de los temas. Además, el hecho de utilizar mensajes de partidos españoles y del Reino Unido permitirá hacer una comparación entre dos países diferentes.

During electoral campaigns, political parties use different communication channels for their messages to reach the voters. Traditionally, the parties' platforms have been the main tool for this. On the other hand, social media is becoming an increasingly important channel in recent times.

The aim of this project is to analyse and compare the political ideology of the messages published by the political parties depending on the media they use and for different topics, among which I will show the results for *equality*.

To accomplish this, a selection of tweets and parties' platforms are used. The chosen parties are the five national parties with most representation in the last elections in Spain and four national parties in the UK.

The project is expected to examine whether there are ideological differences between the messages depending on the media being used for each party and if the ideology gets radicalised. Moreover, analysing similar aspects from Spanish and UK political parties may reflect a broader idea of similarities or differences at a European level.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and justification

This project is important to understand the influence of social media in political communications between parties and voters. Especially, since the spread of fake news and its negative impacts are increasingly undermining the basic pillars of democracy on a global scale, this project aims to study this on the European level.

This is an ongoing problem which is increasingly being studied in more detail, majorly in the USA and UK. However, it has gained little attention in Europe so far. Additionally, European regulations are stricter in context to influence of social media in political campaigns, but complete avoidance of it is neither possible nor desirable.

In reality, the regulations need to be carefully drafted to allow the fair use of social media instead of completely eliminating it because it can be a tool of wider reach for political parties in the technological era that we are living in. Drafting such regulations require a detailed understanding of how social media is being used by political parties, for example, whether to spread their ideologies to a wider audience (especially younger generations) or to manipulate by deceitful messages.

This project is aimed at precisely this aspect, to analyse whether the messages by political parties on social media (Twitter) are in sync with their ideology mentioned on their official platforms.

This project is especially fascinating given the rise of social media in recent years and their usage for political cynicism. The objective is to understand the influence of social media in political communications, and the results of proposed analysis may help to devise strategies to prevent the spread of misinformation.

1.2 Objectives

The main objectives of the project are as follows:

- Analyse whether political ideology of the social media messages is different than that of their official platforms.
- Compare the ideology of political messages between parties from different spectra.
- Analogise the results between countries (Spain and United Kingdom) to obtain a better understanding at the European level.

1.3 Planification

Table 1.1 describes the timeline of the project detailing which tasks will be carried out at which point in time.

Table 1.1: Timeline of the project

Task	Start Date	End Date	Timeline																		
			Month 1				Months 2 and 3									Month 4					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
PAC1: Definició i planificació	21/09/2020	27/09/2020																			
PAC2: Estat de l'art	28/09/2020	18/10/2020																			
PAC3: Disseny i implementació	19/10/2020	20/12/2020																			
Text mining	19/10/2020	25/10/2020																			
Spain	26/10/2020	22/11/2020																			
United Kingdom	23/11/2020	20/12/2020																			
PAC 4: Redacció de la memòria	21/12/2020	03/01/2021																			
PAC5: Presentació i defensa	04/01/2021	10/01/2021																			
Defensa pública	11/01/2021	20/01/2021																			

Chapter 2

State of the art

The first time that an electoral campaign was analysed on a social networking site (SNS) was in 2008 for the presidential campaign of the United States [1]. Other researchers such as Lee [2] had already explored the online political climate during the 2007 political campaign in South Korea, but not on SNS. The candidates were Barack Obama and John McCain. The SNS considered were Myspace (founded in 2003), Facebook (founded in 2004) and YouTube (founded in 2005), and the aim was to examine the relationship between SNS and political cynicism, and the influence of user's background in it. Even though being one of the first of its kind, the authors analysed only one country. Also, their primary goal was to examine the impact of the use of SNS on political cynicism and they did not use NLP techniques.

In Europe, the first use of Twitter to analyse electoral campaigns was during the German elections of 2009 [3] and the Swedish elections of 2010 [4]. The aim of Tumasjan [3] was to examine whether twitter is used for political deliberation, to evaluate whether Twitter messages reflect the current political sentiment and whether they can be used to predict the popularity of parties. The authors examined 104003 political tweets published during one month before the election. The corpus contained the names of the six parties with representation in the German parliament and a selection of relevant politicians based on a weekly survey of popularity. To analyse the political sentiment, they used the relative frequencies of LIWC category word counts and they compared the results with the election programs and the press. The results showed that political sentiments on Twitter messages by parties and candidates do align with the political campaign and programs.

Larsson [4] used 99832 tweets from a time span of a month (August 17th to September 22nd, 2010) before the 2010 Swedish election. The goal of this work was different as compared to previous approaches. Here, instead of categorising the content or use metrics on tweets, the authors identify different user types based on high-end users. For this, the tweets were divided in categories between singleton, directed messages and retweets. The tweets were collected from anonymous citizens and politicians.

In both Tumasjan and Larsson, the authors analyse the use of Twitter only in one country, and they did not compare the results with official programs. Also, they did not use NLP methods.

More recent studies like Deltell [5] analysed the relationship between the impact of profiles of politicians and parties and the results of the election. However, they did not compare different countries, and instead chose the 2012 election in Andalusia (Spain). The authors selected the Twitter profiles of six parties with a higher number of votes in the 2011 general elections in Andalusia and its leaders in the 2012 regional election. The time span was the electoral campaign (from March 9^{th} to March 23^{rd}, 2012). Then, they counted the number of followers for each party on the first day and the last day of the campaign. Once they had the percentages of followers, they compared the results with different surveys and the results of the election. They showed that their method predicts better than the surveys based on difference of number of votes between the first two parties. However, their method was not able to predict the results for the small parties, probably because the followers were not too active on Twitter yet. Also, they did not compare the political ideology in tweets with the official platforms nor did they used NLP techniques.

Other authors have used machine learning techniques in order to extract the political ideology from the tweets. Most of them use a binary classification based on conservative or progressive ideologies. Traditionally, there are three main approaches: sentiment analysis (SA), machine learning and network-based methods.

The analysis by Pla and co-workers [6] is a good example of SA for detecting political ideology on tweets. The authors used the TASS2013 corpus containing 68000 tweets written in Spain between November 2011 and March 2012. They defined a metric to measure the political tendency as a function of the sum of the polarity multiplied by the tendency for each tweet of every user divided by the number of entities. The results showed that the F1 score is the best obtained on this corpus (0.709), although it struggled in identifying the centre (0.431). The metric defined by them is a good contribution to the literature, however, they also did not compare the results with the official platforms and their analysis was restricted to one country (Spain).

Pennacchiotti [7] and Said-Hung [8] used machine learning techniques to identify political ideology. In Pennacchiotti these techniques were used for three different tasks: political ideology detection, ethnicity identification and affinity for a business. The algorithm used was Gradient Boosted Decision Tress (GBDT) and the corpus were the tweets of 14M users active in April 2014. The model's precision and accuracy were high for the political affiliation task (above 0.80), but the recall was not. This may be because the authors did not use other techniques such as link algorithms or n-grams. They also did not compare the results with the official platforms and different countries.

In Said-Hung [8], a corpus of 24900 tweets from the Spanish general election of May 24^{th}, 2015 was used to describe the tweets of the voters and to estimate the significance of the ideological messages. They did not compare the results with other countries or with the official platforms. The procedure was threefold: the messages were pre-processed with term frequency-inverse document frequency (TF-IDF), then four different machine learning algorithms were applied (Naive Bayes, Support Vector Machines, k-Nearest Neighbors and Random Forests) and finally the best model was applied and SPSS was used to put into test the objectives of the authors. The results showed that Random Forest turned out to be the best algorithm, with the precision of 62.6%, which is close to other similar studies. Also, most of the tweets have neutral ideology (92.5%), and among those with a clear ideology, the 88.3% is progressive. One way to improve the results could be by using Word2vec instead of TF-IDF and to pre-process the data with NLP techniques like n-grams detection.

Barberá [9] is a good example of a network-based model used to identify political ideology on twitter messages. The approach of the author was to build a Bayesian model to identify the user's ideology by examining which political actors every user was following. The model was applied in the United States and five European countries. The political actors were chosen among representatives in national-level institutions, political parties, and media outlets and journalists who tweet about politics. They had at least more than five thousand followers in the case of the US, and two thousand in the European countries. The number of political actors per country varied between 118 and 318. Then a list of their followers was obtained, discarding those less active. Then, the author compared the ideal point estimates of the actors' Twitter network with their DW-NOMINATE (voteview.com) scores in the case of the US, or with measures based on surveys of experts in the case of European countries.

For the US, results showed that the method can classify accounts based on the party they belong to and cluster together those belonging to the same party. For Europe, the results showed some discrepancies between tweets and surveys, although in many cases it can be explained through the effect of coalitions and experts' assessment. An improvement in this approach can be found on Gu [10]. The author makes a comparison between different countries, but did not compare the ideology on tweets and official platforms. Also, other approaches such as machine learning or SA were not employed.

So far, all the methods introduced are based on binary classification of political ideology in tweets by all the members of the 113^{th} U.S. congress (2013-2015) and their followers. Also, on Barberá, the heterogeneous types of links on SNS were ignored. For example, an author with a political ideology closer to the left wing can follow political actors from both left and right wings. To solve this issue, authors in Gu (2016) proposed a network-based method that also mention the retweet links in addition to follow links. This approach predicts a series of continuous ideal points on tweets, instead of binary points. The results showed that this method improves baseline methods and is more aligned with human intuitions. The authors improve the network-based method for political ideology detection, but their analysis was focused only on the US and they did not compare the results with the parties' official platforms. Also, they did not use NLP techniques to compare the results.

Finally, more recent studies have used the advantages of neural networks to tackle the problem with a different approach. For example, in Xiao [11], the authors proposed a new network-based model to detect twitter ideology via multi-task and multi-relational embeddings (TIMME). They compared the results with other models such as Graph Neural Networks (GNN) and Heterogeneous Graph Attention Network (HAN). They use four data-sets of tweets from US accounts: one with only politicians, one with politicians and users keen on political affairs, one with users with moderate interest in politics and another one as a union of the other three. They tested the model in follow relation, reply relation, retweet relation, like relation and mention relation. In most of the cases, TIMME performed better than the baseline models. Also, this model is robust to hyper-parameters, stable and scalable. The model was used to find the overall ideology in the United States based on a binary classification (liberal and conservative). Despite the good results of the model, the authors did not compare political ideology in different countries as well as that on Twitter messages with the official parties' platforms.

This project will contribute to the literature by comparing the political ideology in Twitter messages and parties' official platforms. This will provide a better insight about how political parties communicate with the voters through social media (i.e., Twitter) and if they follow the messages given on their official platforms. Additionally, a comparative analysis between two countries will be performed in order to understand if the political communications happen in similar manner in different socio-cultural environments. Finally, different word embedding models will be compared.

Chapter 3

Methodology

In this project, the political ideologies on Twitter will be compared with the electoral programs of a set of parties from Spain and United Kingdom. The objective is to explore whether parties convey the same ideological messages irrespective of their channel of communication (i.e., Twitter and official platform) for a set of topics. In case they do not, it will be examined whether the ideological messages are more extreme on Twitter compared to the manifestos and the reasons behind, based on Barberá (2015).

Party selection

Five political parties from Spain and four from UK have been selected. The selection of parties has been done according to their national representation in each country, respectively. Additionally, only those parties are selected for which the manifestos are available in the Manifesto Project. Specifically, these parties are:

Spain	Partido Socialista Obrero Español (PSOE)
	Partido Popular (PP)
	Vox
	Unidas Podemos (Podemos)
	Ciadadanos
United Kingdom	Conservative Party (Conservatives)
	Labour Party (Labour)
	Liberal Democrats (Liberal)
	Green Party (Green)

Timeline

The timeline of interest for this work will be from 10^{th} of October to 10^{th} of November 2019 for Spain, and from 12^{th} of November to 12^{th} of December for the United Kingdom.

Corpus

To achieve this, two corpus will be used for each party: a corpus of manifestos and a corpus of tweets. The corpus of manifestos has been obtained from The Manifesto Project [12], which provides datasets with manifestos for over 1000 parties from across the globe starting from 1945 until 2020.

In the dataset, the manifestos are divided in quasi-sentences which have a code based on certain keywords, such as “culture”, “education”, “economic growth”, and so on. The list of keywords can be found in Table 3.1. For some of them, the code indicates a deviation towards left or right. The selection of few topics based on keywords shared by all parties in the manifestos will be used for analysis of this project.

The corpus of tweets has been collected collect from those by political leaders of every party over the month before the election day of each country. In the case of Spain, the first and second leader of each province have been chosen. In the case of UK, since there are 650 parliamentary constituencies, 100 constituencies and their leaders have been randomly chosen. The result is 31,350 tweets for Spain and 34,649 tweets for the UK. For some parties, such as Vox and Conservative Party, the number of tweets is not sufficient. This is due to the fact that for the same number of politicians, they have less amount of tweets indicating their lesser use of Twitter during the same time frame.

Corpus pre-processing

Before using the models, each corpus will be pre-processed. For both manifestos and tweets, stopwords and non-alphanumeric characters will be removed, keeping only nouns and adjectives as per the Stanford POS tagger [13].

Models

For each party and corpus, two models, Word2vec [14] and GloVe [15], will be used to find the most similar words to the keywords selected in Section 3. Then, a selection of a set of words will be made based on related keywords in Table 3.1 and idiosyncrasy of the party. The use of these models instead of TF-IDF will offer more in-depth comparisons of messages because they are more context sensitive methods.

Table 3.1: Keywords from The Manifesto Project by party

Keyword	Tendency	PSOE	PP	Vox	UP	C's	CP	Green	LD	LP
Foreign Special Relationships	Positive						✔			
Foreign Special Relationships	Negative						✔			
Military	Positive	✔	✔	✔	✔	✔	✔		✔	✔
Military	Negative	✔	✔	✔	✔	✔	✔		✔	✔
Peace		✔			✔			✔	✔	✔
Internationalism	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
European Community/Union	Positive	✔	✔	✔	✔	✔		✔	✔	✔
Internationalism	Negative			✔		✔				
European Community/Union	Negative				✔		✔	✔		✔
Constitutionalism	Positive		✔	✔		✔	✔		✔	✔
Constitutionalism	Negative		✔	✔		✔	✔		✔	✔
Decentralization		✔	✔		✔		✔	✔	✔	✔
Centralisation		✔	✔	✔		✔				
Governmental and Administrative Efficiency		✔	✔	✔	✔	✔	✔	✔	✔	✔
Corruption		✔		✔	✔	✔	✔	✔	✔	✔
Free Market Economy			✔	✔		✔	✔		✔	
Incentives	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Market Regulation		✔	✔		✔	✔	✔	✔	✔	✔
Economic Planning					✔		✔			
Corporatism/Mixed Economy		✔	✔		✔				✔	✔
Protectionism	Positive	✔	✔		✔		✔		✔	✔
Protectionism	Negative	✔	✔		✔		✔		✔	✔
Economic Goals		✔	✔	✔	✔	✔				
Keynesian Demand Management			✔	✔	✔			✔		
Economic Growth	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Technology and Infrastructure	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Controlled Economy		✔			✔		✔	✔	✔	✔
Nationalisation					✔			✔		✔
Economic Orthodoxy		✔	✔			✔	✔		✔	
Environmental Protection		✔	✔		✔	✔	✔	✔	✔	✔
Culture	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Equality	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Welfare State Expansion		✔	✔	✔	✔	✔	✔	✔	✔	✔
Welfare State Limitation		✔	✔	✔	✔	✔	✔	✔	✔	✔
Education Expansion		✔	✔	✔	✔	✔	✔	✔	✔	✔
Education Limitation		✔	✔	✔	✔	✔	✔	✔	✔	✔
Traditional Morality	Positive		✔	✔		✔	✔	✔		
Traditional Morality	Negative		✔	✔		✔	✔	✔		
Labour Groups	Positive	✔	✔	✔	✔	✔	✔	✔	✔	✔
Labour Groups	Negative	✔	✔	✔	✔	✔	✔	✔	✔	✔
Middle Class and Professional Groups			✔			✔				
Underprivileged Minority Groups		✔								
Non-economic Demographic Groups		✔	✔	✔			✔	✔	✔	✔

Table 3.2: Weights Global Political Survey, 2019

Party	Weight
Podemos	-0.75
PSOE	-0.48
Ciudadanos	-0.13
PP	0.57
Vox	0.92
Green	-0.64
Liberal	-0.55
Labour	-0.51
CP	0.42

Tendency calculation

To calculate the the political tendencies, a measure similar to the one in [6] for the manifestos 3.1 and tweets 3.2 is used. The formulas are:

$$Political\_Tendency(M_i) = \frac{\sum_{j=1...|P_j|} |W_{ij}| \cdot Weight(P_j)}{\sum_{j=1...|P_j|} |W_{ij}|} \tag{3.1}$$

$$Political\_Tendency(T_i) = \frac{\sum_{j=1...|P_j|} |W_{ij}| \cdot Weight(P_j)}{\sum_{j=1...|P_j|} |W_{ij}|} \tag{3.2}$$

In Eq. 3.1, M is the manifesto of a party i, and P are the parties in the same country as i. W are the most similar words from each one of those parties as explained in 3. *Weight* is a weight given to each party according to the the Global Political Survey, 2019 (GPS19) by Harvard University [16], as seen in Table 3.2.

The weights have been taken from the variable V6 from the [16]. This variable refers to the ideological values based on the degree of social liberalism and conservatism. In other words, it indicates if a party is more liberal or more conservative in their social values. The values' domain is $[0, 10]$, so the values have been standardised between -1 and 1 in order to use them as a weight.

Eq. 3.2 is similar to that for the manifestos, but used for a tweet T. In this case, the final political tendency of all tweets of a party is the average of all their tendencies.

The measures of political tendencies for each party to be used, will be obtained using above equations, and their manifestos will be compared to Twitter. This will be done separately for each of the topics chosen earlier, to understand how each party is deviating from their manifestos in different aspects if the differences are present.

This comparison will demonstrate which parties are conveying the same ideology on Twitter as in their official manifestos and which ones are not. For the parties that are deviating on Twitter by a clear margin, the words responsible to push their tendencies away from that of their manifestos will be investigated.

Also, a description of messages with different techniques will be made. Following this, a set of word clouds for each topic will be created in order to identify the most used words for each party and each model. Finally, word clouds will be created to demonstrate the deviations of parties on Twitter.

Model evaluation and results

Each model will be evaluated on each keyword and each corpus by finding the difference between the tendency and the standardised weights. It will be shown that the results of those models and keywords which difference between tendency and weights is smaller across parties.

Novelty of the approach

The comparison between these two approaches using the above-mentioned specific techniques has not been performed before to the best of my knowledge, especially involving Twitter and official platforms.

Chapter 4

Results and Discussion

In this section, the results of experiments conducted on Spanish and UK parties are shown for the keyword *equality* and the model Word2vec.

For each topic and model, words responsible for the deviation between the two platforms are explored if they are present, such as *education*, *international*, and *economy*, and *equality*. The results will be shown only of the experiments for the word *equality*.

Table 4.1: Comparing Word2vec and GloVe for "igualdad"/"equality"

Igualdad/Equality												
Country	Party	Weights	W2v					GloVe				
			Tendency manifesto	Tendency twitter	Δ weight-manifesto	Δ weight-twitter	Δ man-tw	Tendency manifesto	Tendency twitter	Δ weight-manifesto	Δ weight-twitter	Δ man-tw
Spain	UP	-1,02	-0,42	-0,08	0,6	0,94	0,34	-1,42	-0,85	0,4	0,17	0,57
	PSOE	-0,58	-1,13	0,05	0,55	0,63	1,18	0,31	-1,29	0,89	0,71	1,6
	PP	1,13	0,16	0,1	0,97	1,03	0,06	0,4	0,47	0,73	0,66	0,07
	C's	0	-0,19	0,73	0,19	0,73	0,92	-0,51	0,91	0,51	0,91	1,42
	Vox	1,72	1,58	0,73	0,14	0,99	0,85	1,22	0,76	0,5	0,96	0,46
United Kingdom	LP	-0,62	-0,18	-0,08	0,44	0,54	0,1	-0,44	0,71	0,18	1,33	1,15
	LD	-0,7	-0,47	0,89	0,23	1,59	1,36	-0,29	0	0,41	0,7	0,29
	GP	-0,84	-0,8	-1,37	0.04	0,53	0,57	-0,74	-1,41	0,1	0,57	0,67
	CP	0,9	1,45	0,57	0,55	0,33	0,88	1,47	0,71	0,57	0,19	0,76

Between the two models, both Word2vec and GloVe have performed in a similar way, as shown in Table 4.1. Word2vec has performed better for manifesto since the tendencies of the manifestos are found to be closer to the original weights from [16], whereas GloVe has performed slightly better for Twitter. Since, Word2Vec is significantly better than GloVe for manifestos and only slightly behind GloVe for Twitter, only the results of the experiments with Word2vec will be shown for this project.

4.1 Spain

From Spain, five national parties have been selected, namely, PSOE, PP, Vox, Podemos, and Ciudadanos.

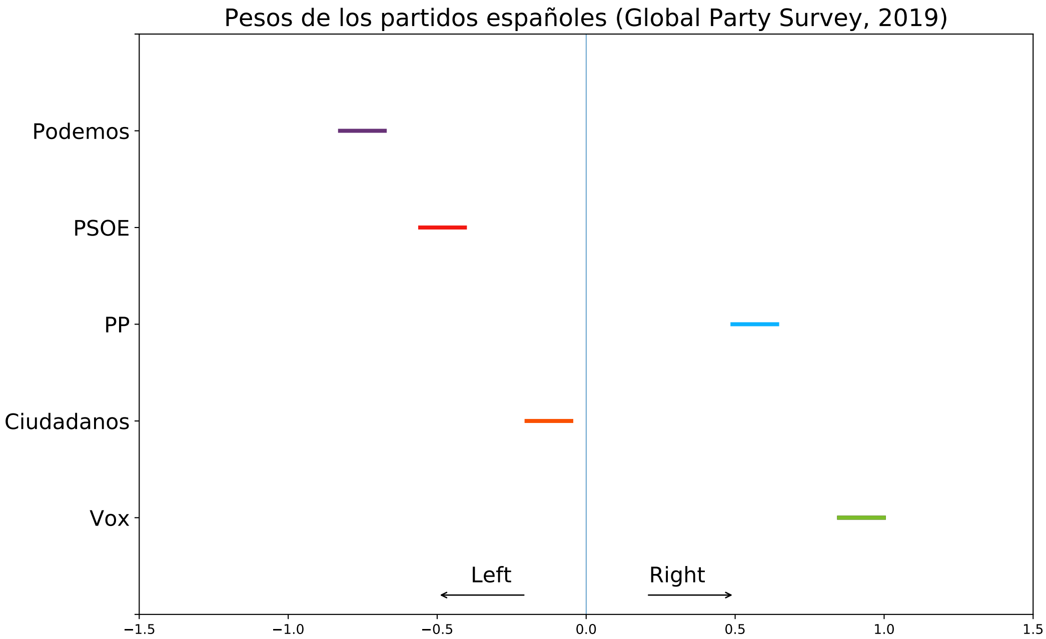


Figure 4.1: Ideological weights for Spanish parties.

For each party, Word2vec model is applied to both manifesto and Twitter corpus to find out the most similar words. A few words are selected among those not shared with other parties. These words are related to the topics of Table 3.1.

In this section, comparison of political tendencies of each of the five parties in Spain for the topic *igualdad* is shown. The goal is to check which parties deviate more in Twitter messages from their official manifestos for this topics.

Additionally, the influence of certain words that contribute to their deviation on Twitter (whether left or right leaning) from manifestos will be demonstrated using word clouds.

In Figure 4.2, the political tendencies of all the parties in both manifesto and tweets are described using the model Word2vec for the topic *igualdad*. Observe that all parties show slight deviations on Twitter except for PP. Specifically, on Twitter, Podemos and Vox deviate to the left, and PSOE and Ciudadanos, to the right. It is also worth noting that, for manifestos, PSOE's tendency is more to the left than that of Podemos.

Since Podemos and Vox represent two ends of the political spectrum (left and right, respectively) according to the weights (see Fig. 4.1), they are selected for further analysis to explore the reasons behind their deviations on Twitter. By comparing the words appearing in their tweets and manifestos, it can be demonstrated which words contribute to tilting their tendencies in one direction.

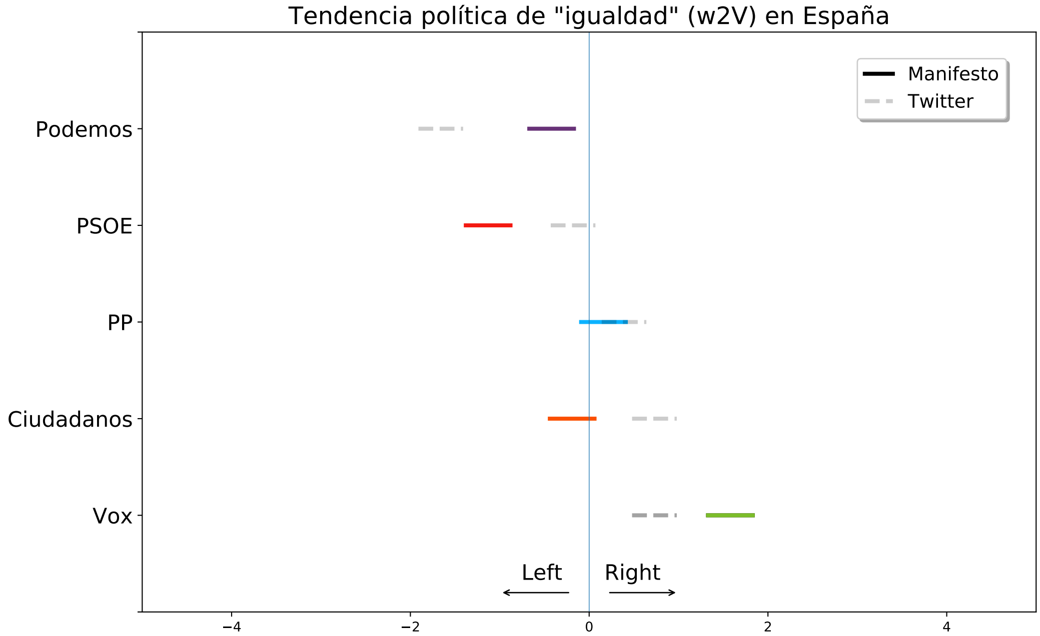
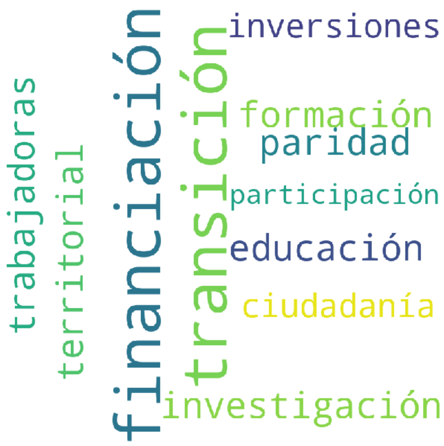
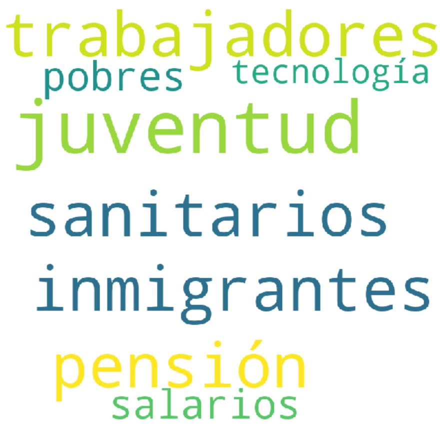


Figure 4.2: Comparison between political tendency in manifestos and tweets for Spanish parties for *igualdad* using Word2vec.



(a) Manifesto



(b) Twitter

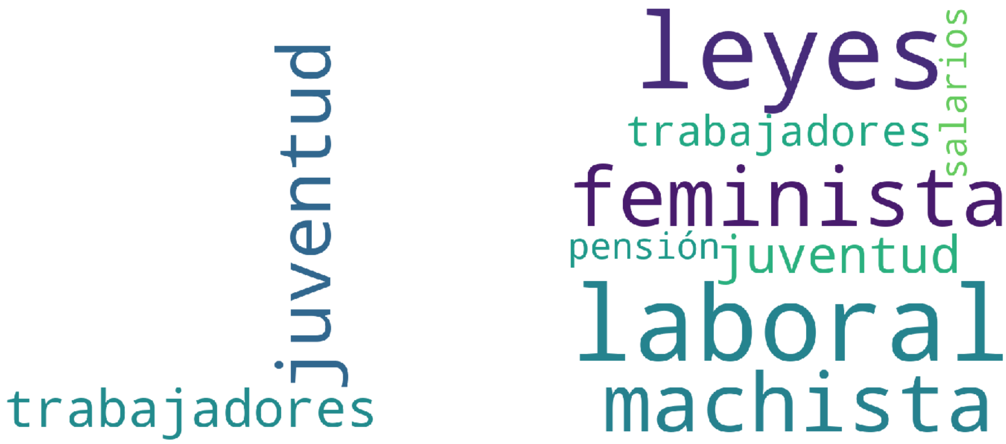
Figure 4.3: Wordclouds of Podemos for *igualdad* obtained using Word2vec.

In Fig. 4.3, the wordclouds of Podemos are shown. On manifesto, Podemos talks about *transición*, *educación*, *investigación*, whereas on Twitter, they talk about other topics, such as, *immigrantes*, *pensión*, *salarios*, and *juventud*, among others.

(a) Manifesto (b) Twitter

Figure 4.4: Wordclouds of Vox for "igualdad" obtained using Word2vec.

In Fig. 4.4, the wordclouds of Vox are shown. On manifesto, Vox talk about *delito*, *terrorismo*, *immigrantes*, *fronteras*, *mafia*, whereas on Twitter, they talk about *marxismo*, *separatismo*, and *reconciliación*, among others.



(a) Vox (b) Podemos

Figure 4.5: Wordclouds demonstrating left leaning tendencies of Vox and Podemos for *igualdad* using Word2vec.

On the other hand, the reason for Podemos' deviation on twitter to the left can be associated to words generally used by left parties. Figure 4.5b shows that they use words like *feminista* and *machista* which are used by PSOE and are responsible for pushing Podemos to further left compared to their manifesto.

In the case of Vox, the results are not conclusive because the corpus is smaller than the other parties. The model only found two words, *juventud* and *trabajadores*, which are insufficient to say with certainty that Vox leans towards left in reality.

4.2 United Kingdom

From the United Kingdom, four national parties are selected, namely, Conservative Party (CP), Green Party (Green), Liberal Democrats (Liberal), and Labour Party (Labour). Their weights from the GPS19 are shown in Figure 4.6. It is interesting to note that, unlike Spain, all UK parties are on the left ideological spectrum except for Conservative Party.

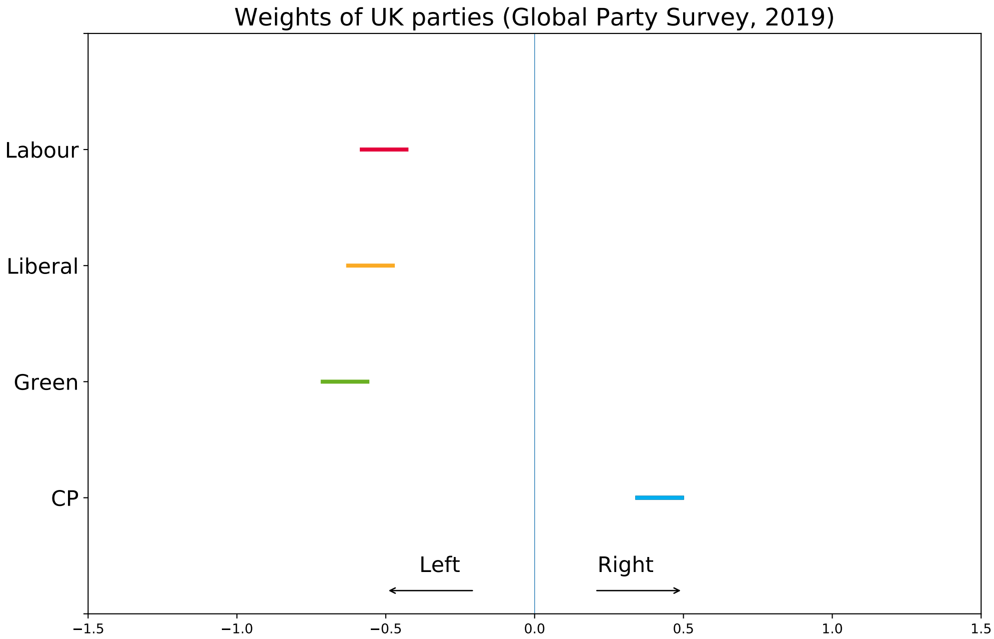


Figure 4.6: Ideological weights for UK parties.

In this section, the political tendencies are compared for each of the four parties in the United Kingdom (UK) for the topic *equality*.

In Figure 4.7, the political tendencies of UK parties for both platforms, manifesto and Twitter, are described using the model Word2vec for the topic *equality*.

Observe that Labour does not show any notable deviations whereas Green leans a little to the left. The other two parties, however, show significant deviations with CP leaning to left and Liberal to right.

Further analysis of CP and Liberal is carried out to capture the words behind their deviation. This will lead to some understanding of why these two parties show larger differences on two platforms as compared to other parties.

Figure 4.8 shows wordclouds representing selected words from manifesto and Twitter for Liberal party. Notice that, their use of words like *zerocarbon*, *youth* and *welfare* in their manifesto, while *freedom*, *investment*, *expansion*, and *workforce* on Twitter.

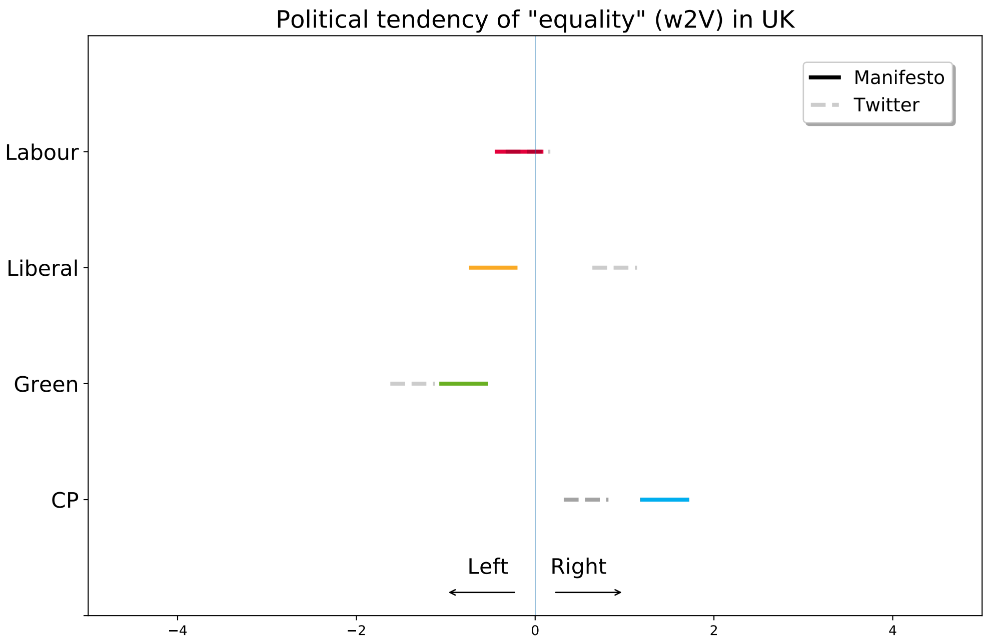
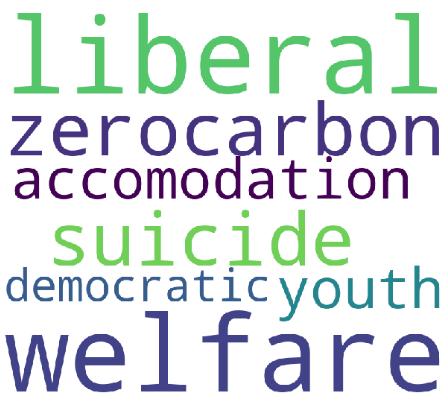
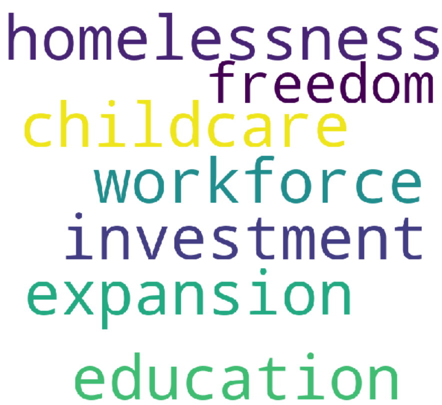


Figure 4.7: Comparison between political tendency in manifestos and tweets for UK parties for *equality* using Word2vec.



(a) Manifesto



(b) Twitter

Figure 4.8: Word clouds of Liberal for *equality* obtained using Word2vec.

On the other hand, Figure 4.9 shows that, CP talks about *immigration*, *independence*, *prosperity*, and *commonwealth* in their manifesto, ideas that seem to be related to Brexit. However, on Twitter, they talk about *poor*, *violence*, *constitution*, and *safety*, so that their messages appear more populist.

(a) Manifesto

(b) Twitter

Figure 4.9: Word clouds of CP for *equality* obtained using Word2vec.

(a) Conservative

(b) Liberal

Figure 4.10: Wordclouds demonstrating left leaning tendency of CP and right leaning for Liberal for *equality* using Word2vec.

Figure 4.10 shows the words that make CP's tendency on Twitter deviate to the left and that of liberal to the right.

Note that, for Liberal the deviation is because of words such as *british*, *safety*, and *church*, more used by Conservative Party. This might have helped push their tendency to appear more right.

For CP, the word *sustainability* is found, a term largely used by Green Party. Hence, on manifesto, they seem to care about climate change and topics related to youth, making their tendency to be left leaning, whereas on Twitter they appear to talk more about topics traditionally associated to right. Although, this conclusion should be taken with a grain of salt since only one word has been found to demonstrate this. A more robust method of word selection may be necessary to improve the reliability of this approach, which is outside the scope of this project at the moment. It will be an interesting aspect for the future work in this field to explore.

Chapter 5

Conclusions

In this project, the political tendencies of a set of parties from Spain and the UK have been compared between manifestos and tweets. The tendency has been calculated in relation to a number of topics and using Word2vec and GloVe word embedding models.

The results have shown that, for the topic *equality*, in the case of Spain, most parties deviate on Twitter as compared to their manifestos. On the contrary, in the UK, the parties' tendency seems to be more similar between manifestos and Twitter.

For example, in Spain, Podemos's tendency in Twitter is more to the left than that of manifesto. This is because on Twitter, they use more words from the other Spanish left party, PSOE. For PSOE and Ciudadanos, their tendencies on Twitter deviate to the right because on Twitter, they are using words from right wing parties such as Vox or PP.

It is also interesting to note that Vox uses words such as *terrorismo*, *delito*, and *inmigrantes*, whereas on Twitter they talk about *adoctrinamiento*, *marxismo*, and *separatismo*. Therefore, on manifesto they look more concerned about crime and terrorism, whereas on Twitter they appear more bothered about pursuit of independence by Spanish regions.

For UK, the only clear case where tendencies are different are the Liberal Democrats, whose tendency on manifesto leans on the left, whereas the tendency on Twitter leans on the right, mainly because of words used by Conservative party such as *british* and *safety*. The rest of parties' tendencies remain quite similar in general.

Overall, it could be observed that Spanish parties have a higher tendency to be influenced by the speeches of other parties than in the UK. It is also interesting to point out that Spanish parties' words on both manifesto and Twitter are more extreme than those used by UK parties (i.e. *terrorismo*, *adoctrinamiento*).

Chapter 6

Future work

There are many interesting observations which could benefit from a more detailed analysis. There are several directions in which the work of this project could be extended.

For example, one interesting analysis could be to use official party accounts instead of the Twitter accounts of party leaders for comparison with manifesto. This would eliminate the personal biases of party leaders, because personal Twitter accounts are bound to have personal opinions of the people involved, which may or may not strictly align with their parties' ideologies.

Another interesting aspect would be time dependent analysis. More specifically, to examine the tweets on a daily or monthly basis and to check if certain events (local or global) influence their behaviour and how.

Additionally, newer models like BERT could be used to find the tendencies and compare the results with Word2vec and GloVe. Being a bidirectional model, BERT could eliminate the contextual biases of words that are likely to be noisy owing to their neutral nature.

Furthermore, increasing the set of topics to analyse the parties and their ideologies would reduce errors and make the analysis more robust. Analysis can also be done on topics that are not necessarily shared by all the parties of interest to reveal their unique tendencies if they exist.

Finally, comparing more countries at a European level and/or to include other countries from different continents such as Asia. This would extend our understanding of how politics works in different socio-cultural as well as different socio-economical environments.

Bibliography

[1] G. Hanson, P. Haridakis, A. Cunningham, R. Sharma, and J. Ponder, "The 2008 presidential campaign: Political cynicism in the age of facebook, myspace, and youtube," *Mass Communication and Society - MASS COMMUN SOC*, vol. 13, pp. 584–607, 10 2010.

[2] Y.-O. Lee and H. W. Park, "The reconfiguration of e-campaign practices in korea: A case study of the presidential primaries of 2007," *International Sociology*, vol. 25, no. 1, pp. 29–53, 2010.

[3] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," vol. 10, 01 2010.

[4] A. O. Larsson and H. Moe, "Studying political microblogging: Twitter users in the 2010 swedish election campaign," *New media & society*, vol. 14, no. 5, pp. 729–747, 2012.

[5] L. Deltell, F. Claes, and J. M. Osteso, "Predicción de tendencia política por twitter: Elecciones andaluzas 2012," *Ámbitos. Revista internacional de comunicación*, no. 22, 2013.

[6] F. Pla and L.-F. Hurtado, "Political tendency identification in twitter using sentiment analysis techniques," in *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical Papers*, pp. 183–192, 2014.

[7] M. Pennacchiotti and A.-M. Popescu, "A machine learning approach to twitter user classification," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, 2011.

[8] E. M. Said-Hung, R. C. Prati, and A. Cancino-Borbón, "La orientación ideológica de los mensajes publicados en twitter durante el 24m en españa," *Palabra Clave*, vol. 20, no. 1, pp. 213–238, 2017.

[9] P. Barberá, "Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data," *Political analysis*, vol. 23, no. 1, pp. 76–91, 2015.

[10] Y. Gu, T. Chen, Y. Sun, and B. Wang, "Ideology detection for twitter users with heterogeneous types of links," *arXiv preprint arXiv:1612.08207*, 2016.

[11] Z. Xiao, W. Song, H. Xu, Z. Ren, and Y. Sun, "Timme: Twitter ideology-detection via multi-task multi-relational embedding," *arXiv preprint arXiv:2006.01321*, 2020.

[12] A. Volkens, T. Burst, W. Krause, P. Lehmann, T. Matthieß, N. Merz, S. Regel, B. Weßels, and L. Zehnter, "The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2020b," 2020.

[13] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology-volume 1*, pp. 173–180, Association for Computational Linguistics, 2003.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[15] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[16] P. Norris, "Global Party Survey, 2019," 2020.